

REMARKS

Claims 1-36 were pending in this application. In a Final Office Action dated August 24th, 2007, claims 1-36 were rejected.

Applicants are amending claims 1, 3, 6, 8, 12, 13, 24 and 36 in this Amendment and Response. These amendments have been made to clarify the claimed subject matter and their entry is respectfully requested. Claims 37-54 are newly presented.

In view of the Amendments herein and the Remarks that follow, Applicants respectfully request that the Examiner reconsider all outstanding objections and rejections, and withdraw them.

Summary of Interview

Applicant's representative thanks Examiner Paras Shah and Supervisory Examiner Patrick Edouard for their time in conducting an interview on October 4th, 2007. The relevant portions of this discussion are summarized herein in accordance with MPEP §713.04.

During the interview, Applicant's representative and the Examiners discussed the rejection of claims 1, 3-5, 13 and 15-23 under 35 USC 112. Applicant's representative and the Examiners further discussed the rejection of independent claims 1, 6, 12, 13, 24 and 36 under 35 USC 103(a). During this discussion, argument was put forth that the combination of references failed to teach each and every element of the claimed invention. Specifically, substantial explanation was given as to why Su failed to disclose an iterative method. These arguments and explanations are summarized below. The Applicant's representative also agreed to make amendments to the claims to clarify the claimed subject matter and facilitate prosecution.

Response to Rejection Under 35 USC § 112, Paragraph 2

In the 7th, 8th and 9th paragraphs of the Office Action, the Examiner rejects claims 1, 3-5, 13 and 15-23 under 35 USC 112, Paragraph 2 as allegedly failing to point out and distinctly claim the subject matter which Applicants regard as invention. Specifically, the Examiner asserts that the phrase “configured to” renders the claims indefinite since it suggests optional language.

Independent claims 1 and 13 have been amended to recite “executable to” in place of “configured to”. In the amended claims, the term “executable” is used to denote that the components of the systems in claims 1 and 13 deterministically perform a set of functions when executed or run. As the set of functions are deterministically run, Applicants submit that claims 1 and 13 do not recite optional language. Based on these amendments, Applicants submit that the claims point out and distinctly claim the subject matter which Applicants regard as the invention.

Response to Rejection Under 35 USC 103(a)

In the 10th and 11th paragraphs of the Office Action, the Examiner rejects claims 1, 3, 6, 8, 11-13, 20-24 and 31-36 under 35 USC 103(a) as allegedly being unpatentable over Su et al. (In Proceedings of the 32nd Annual Meeting of the Association for Computation Linguistics, 1994) in view of Jurafsky et al. (Speech and Language Processing: An Introduction to Natural Language Processing). This rejection is respectfully traversed.

The claimed invention is directed systems, methods and apparatus which use a vocabulary comprising tokens to iteratively identify compounds having a plurality of lengths within the text corpus. At each iteration, a set of *n*-grams having a same length is identified and *n*-grams are added to the vocabulary. At least part of the vocabulary is rebuilt at each iteration based on the added *n*-grams.

To facilitate prosecution, the independent claims have been amended to clarify that the claimed limitations are directed to an iterative method wherein elements such as the length of n-grams and the vocabulary are modified at each iteration. Specifically, independent claims 1, 6, 2, 13, 24 and 36 have been amended to recite elements similar to:

iteratively identifying compounds having a plurality of lengths within the text corpus and rebuilding at least part of the vocabulary based on the identified compounds having the plurality of lengths, each compound comprising a plurality of tokens

Su does not disclose these features. The system in Su is modeled as a two-class classification problem wherein the classifier is trained on features such as mutual information calculated from a training corpus labeled using a fixed sizes of n-grams. Su discloses a comparison of results obtained from using a classifier trained using bigrams and trigrams.

Specifically, Su does not disclose “iteratively identifying compounds having a plurality of lengths within the text corpus”. In his analysis and during the telephone interview, the Examiner supported the rejection of this element by citing to a portion of Su which discloses building bigram and trigram classifiers. The bigram and trigram classifiers in Su are compared to evaluate the accuracy (*see* Abstract) and the distribution statistics (See Tables 1 and 2) of the method outlined in Su. In Su, the bigram and trigram classifiers are built independently, producing separate classifiers which either identify bigrams or trigrams. Therefore, Su **teaches away** from the claimed invention of “identifying compounds **having a plurality of lengths** within the text corpus” as Su is limited to the identification of compounds of only **one length per classification model**. There is nothing in Su to suggest or even hint at iteratively combining the 2-gram and 3-gram classifiers for “iteratively identifying compounds having a plurality of lengths within the text corpus”.

Su further fails to provide a construction which supports both “a vocabulary comprising tokens extracted from a text corpus” and “rebuilding at least part of the vocabulary based on the identified compounds having a plurality of lengths”. In his rejection, the Examiner cites a portion of Su which discloses the generation of a training corpus using to construct a classifier. As discussed above, Su discloses creating either a bigram classifier or a trigram classifier and therefore does not teach rebuilding at least part of the training corpus “based on the identified compounds having a plurality of lengths”.

Jurafsky does not remedy the deficiencies of Su. Jurafsky is a textbook which outlines standard techniques in speech and language processing. Jurafsky merely discloses “backoff”, an interpolation technique used to calculate likelihood of n-grams based on lesser order n-grams. Jurafsky does not teach or suggest compound identification or a “a vocabulary comprising tokens extracted from a text corpus”. Accordingly, Jurafsky fails to disclose or suggest “iteratively identifying compounds having a plurality of lengths within the text corpus” and “rebuilding at least part of the vocabulary based on the identified compounds having a plurality of lengths”.

Based on at least the above, Applicant’s submit that independent claims 1, 6, 12, 13, 24 and 36 are patentably distinguishable over Su and Jurafsky, alone or in any combination. Additionally, the dependent claims recite features not disclosed by the cited art.

On the basis of the above, Applicants respectfully submit that the pending claims are patentable over the cited art. The early allowance of all claims herein is requested. If the Examiner believes that direct contact with the Applicants’ attorney will advance the prosecution of this case, the Examiner is encouraged to contact the undersigned as indicated below.

Respectfully Submitted,

Franz, et al.

Date: December 26, 2007

By: /Brian Hoffman/
Brian M. Hoffman, Reg. No. 39,713
Attorney for Applicant
Fenwick & West LLP
801 California Street
Mountain View, CA 94041
Tel.: (415) 875-2484
Fax: (415) 281-1350